```
                RAW SEQUENCE LISTING              DATE: 05/09/2000
                PATENT APPLICATION:  US/09/458,366   TIME: 11:12:47

                Input Set : A:\Sequence.txt
                Output Set: N:\CRF3\05092000\I458366.raw
```

ENTERED

```
  4 <110> APPLICANT: Evans, Ronald M.
  5        Blumberg, Bruce
  7 <120> TITLE OF INVENTION: NOVEL STEROID-ACTIVATED NUCLEAR
  8        RECEPTORS AND USES THEREFOR
 10 <130> FILE REFERENCE: SALK2270-2
 12 <140> CURRENT APPLICATION NUMBER: 09/458,366
 13 <141> CURRENT FILING DATE: 1999-12-09
 15 <150> PRIOR APPLICATION NUMBER: 09/005,286
 16 <151> PRIOR FILING DATE: 1998-01-09
 18 <160> NUMBER OF SEQ ID NOS: 39
 20 <170> SOFTWARE: FastSEQ for Windows Version 4.0
 22 <210> SEQ ID NO: 1
 23 <211> LENGTH: 2068
 24 <212> TYPE: DNA
 25 <213> ORGANISM: Homo sapiens
 27 <220> FEATURE:
 28 <221> NAME/KEY: CDS
 29 <222> LOCATION: (583)...(1884)
 31 <223> OTHER INFORMATION: N is selected from A, C, T/U or G
 33 <400> SEQUENCE: 1
 34 ggcacgagga gatctaggtt caaattaatg ttgcccctag tggtaaagga cagagaccct    60
 35 cagactgatg aaatgcgctc agaattactt agacaaagcg gatatttgcc actctcttcc   120
 36 ccttttcctg tgtttttgta gtgaagagac ctgaaagaaa aaagtaggga gaacataatg   180
 37 agaacaaata cggtaatctc ttcatttgct agttcaagtg ctggacttgg gacttaggag   240
 38 gggcaatgga gccgcttagt gcctacatct gacttggact gaaatatagg tgagagacaa   300
 39 gattgtctca tatccgggga aatcataacc tatgactagg acgggaagag gaagcactgc   360
 40 ctttacttca gtgggaatct cggcctcagc ctgcaagcca agtgttcaca gtgagaaaag   420
 41 caagagaata agctaatact cctgtcctga acaaggcagc ggctccttgg taaagctact   480
 42 ccttgatcga tcctttgcac cggattgttc aaagtggacc ccaggggaga agtcggagca   540
 43 aagaacttac caccaagcag tccaagaggc ccagaagcaa ac ctg gag gtg aga      594
 44                                                   Leu Glu Val Arg
 45                                                     1
 47 ccc aaa gaa agc tgg aac cat gct gac ttt gta cac tgt gag gac aca     642
 48 Pro Lys Glu Ser Trp Asn His Ala Asp Phe Val His Cys Glu Asp Thr
 49  5               10                  15                  20
 51 gag tct gtt cct gga aag ccc agt gtc aac gca gat gag gaa gtc gga     690
 52 Glu Ser Val Pro Gly Lys Pro Ser Val Asn Ala Asp Glu Glu Val Gly
 53              25                  30                  35
 55 ggt ccc caa atc tgc cgt gta tgt ggg gac aag gcc act ggc tat cac    738
 56 Gly Pro Gln Ile Cys Arg Val Cys Gly Asp Lys Ala Thr Gly Tyr His
 57              40                  45                  50
 59 ttc aat gtc atg aca tgt gaa gga tgc aag ggc ttt ttc agg agg gcc    786
 60 Phe Asn Val Met Thr Cys Glu Gly Cys Lys Gly Phe Phe Arg Arg Ala
 61          55                  60                  65
 63 atg aaa cgc aac gcc cgg ctg agg tgc ccc ttc cgg aag ggc gcc tgc    834
 64 Met Lys Arg Asn Ala Arg Leu Arg Cys Pro Phe Arg Lys Gly Ala Cys
 65      70                  75                  80
```

```
 67 gag atc acc cgg aag acc cgg cga cag tgc cag gcc tgc cgc ctg cgc      882
 68 Glu Ile Thr Arg Lys Thr Arg Arg Gln Cys Gln Ala Cys Arg Leu Arg
 69  85                  90                  95                 100
 71 aag tgc ctg gag agc ggc atg aag aag gag atg atc atg tcc gac gag      930
 72 Lys Cys Leu Glu Ser Gly Met Lys Lys Glu Met Ile Met Ser Asp Glu
 73                 105                 110                 115
 75 gcc gtg gag gag agg cgg gcc ttg atc aag cgg aag aaa agt gaa cgg      978
 76 Ala Val Glu Glu Arg Arg Ala Leu Ile Lys Arg Lys Lys Ser Glu Arg
 77                 120                 125                 130
 79 aca ggg act cag cca ctg gga gtg cag ggg ctg aca gag gag cag cgg     1026
 80 Thr Gly Thr Gln Pro Leu Gly Val Gln Gly Leu Thr Glu Glu Gln Arg
 81                 135                 140                 145
 83 atg atg atc agg gag ctg atg gac gct cag atg aaa acc ttt gac act     1074
 84 Met Met Ile Arg Glu Leu Met Asp Ala Gln Met Lys Thr Phe Asp Thr
 85        150                 155                 160
 87 acc ttc tcc cat ttc aag aat ttc cgg ctg cca ggg gtg ctt agc agt     1122
 88 Thr Phe Ser His Phe Lys Asn Phe Arg Leu Pro Gly Val Leu Ser Ser
 89 165                 170                 175                 180
 91 ggc tgc gag ttg cca gag cct ctg cag gcc cca tcg agg gaa gaa gct     1170
 92 Gly Cys Glu Leu Pro Glu Pro Leu Gln Ala Pro Ser Arg Glu Glu Ala
 93                 185                .190                 195
 95 gcc aag tgg agc cag gtc cgg aaa gat ctg tgc tct ttg aag gtc tct     1218
 96 Ala Lys Trp Ser Gln Val Arg Lys Asp Leu Cys Ser Leu Lys Val Ser
 97                 200                 205                 210
 99 ctg caa gct gcg ggg gga gga tgg cag tgt ctg gaa cta caa acn ccc    `1266
100 Leu Gln Ala Ala Gly Gly Gly Trp Gln Cys Leu Glu Leu Gln Xaa Pro
101                 215                 220                 225
103 agc cga cag tgg cgg aaa gag atc ttc tcc ctg ctg ccc cac atg gct     1314
104 Ser Arg Gln Trp Arg Lys Glu Ile Phe Ser Leu Leu Pro His Met Ala
105        230                 235                 240
107 gac atg tca acc tac atg ttc aaa ggc atc atc agc ttt gcc aaa gtc     1362
108 Asp Met Ser Thr Tyr Met Phe Lys Gly Ile Ile Ser Phe Ala Lys Val
109 245                 250                 255                 260
111 atc tcc tac ttc agg gac ttg ccc atc gag gac cag atc tcc ctg ctg     1410
112 Ile Ser Tyr Phe Arg Asp Leu Pro Ile Glu Asp Gln Ile Ser Leu Leu
113                 265                 270                 275
115 aag ggg gcc gct ttc gag ctg tgt caa ctg aga ttc aac aca gtg ttc     1458
116 Lys Gly Ala Ala Phe Glu Leu Cys Gln Leu Arg Phe Asn Thr Val Phe
117                 280                 285                 290
119 aac gcg gag act gga,acc tgg gag tgt ggc cgg ctg tcc tac tgc ttg     1506
120 Asn Ala Glu Thr Gly Thr Trp Glu Cys Gly Arg Leu Ser Tyr Cys Leu
121                 295                 300                 305
123 gaa gac act gca ggt ggc ttc cag caa ctt cta ctg gag ccc atg ctg     1554
124 Glu Asp Thr Ala Gly Gly Phe Gln Gln Leu Leu Leu Glu Pro Met Leu
125        310                 315                 320
127 aaa ttc cac tac atg ctg aag aag ctg cag ctg cat gag gag gag tat     1602
128 Lys Phe His Tyr Met Leu Lys Lys Leu Gln Leu His Glu Glu Glu Tyr
129 325                 330                 335                 340
131 gtg ctg atg cag gcc atc tcc ctc ttc tcc cca gac cgc cca ggt gtg    `1650
```

W--> 99 / 100 (marked in left margin)

RAW SEQUENCE LISTING   DATE: 05/09/2000
PATENT APPLICATION: US/09/458,366   TIME: 11:12:47

Input Set : A:\Sequence.txt
Output Set: N:\CRF3\05092000\I458366.raw

```
132 Val Leu Met Gln Ala Ile Ser Leu Phe Ser Pro Asp Arg Pro Gly Val
133               345                 350                 355
135 ctg cag cac cgc gtg gtg gac cag ctg cag gag caa ttc gcc att act      1698
136 Leu Gln His Arg Val Val Asp Gln Leu Gln Glu Gln Phe Ala Ile Thr
137               360                 365                 370
139 ctg aag tcc tac att gaa tgc aat cgg ccc cag cct gct cat agg ttc      1746
140 Leu Lys Ser Tyr Ile Glu Cys Asn Arg Pro Gln Pro Ala His Arg Phe
141               375                 380                 385
143 ttg ttc ctg aag atc atg gct atg ctc acc gag ctc cgc agc atc aat      1794
144 Leu Phe Leu Lys Ile Met Ala Met Leu Thr Glu Leu Arg Ser Ile Asn
145         390                 395                 400
147 gct cag cac acc cag cgg ctg ctg cgc atc cag gac ata cac ccc ttt      1842
148 Ala Gln His Thr Gln Arg Leu Leu Arg Ile Gln Asp Ile His Pro Phe
149 405                 410                 415                 420
151 gct acg ccc ctc atg cag gag ttg ttc ggc atc aca ggt agc              1884
152 Ala Thr Pro Leu Met Gln Glu Leu Phe Gly Ile Thr Gly Ser
153                     425                 430
155 tgagcggctg ccttgggtga caccttcgag aggcagccag acccagagcc ctctgagccg     1944
156 gcactcccgg gccaagacag atggacactg ccaagagccg acaatgccct gctggcctgt     2004
157 ctccctaggg aattcctgct atgacagctg gctagcattc ctcaggaagg acatggggtg     2064
158 cccc                                                                  2068
160 <210> SEQ ID NO: 2
161 <211> LENGTH: 434
162 <212> TYPE: PRT
163 <213> ORGANISM: Homo sapiens
165 <220> FEATURE:
166 <223> OTHER INFORMATION: Xaa is threonine
168 <400> SEQUENCE: 2
169 Leu Glu Val Arg Pro Lys Glu Ser Trp Asn His Ala Asp Phe Val His
170 1               5                   10                  15
171 Cys Glu Asp Thr Glu Ser Val Pro Gly Lys Pro Ser Val Asn Ala Asp
172               20                  25                  30
173 Glu Glu Val Gly Gly Pro Gln Ile Cys Arg Val Cys Gly Asp Lys Ala
174               35                  40                  45
175 Thr Gly Tyr His Phe Asn Val Met Thr Cys Glu Gly Cys Lys Gly Phe
176         50                  55                  60
177 Phe Arg Arg Ala Met Lys Arg Asn Ala Arg Leu Arg Cys Pro Phe Arg
178 65                  70                  75                  80
179 Lys Gly Ala Cys Glu Ile Thr Arg Lys Thr Arg Arg Gln Cys Gln Ala
180                 85                  90                  95
181 Cys Arg Leu Arg Lys Cys Leu Glu Ser Gly Met Lys Lys Glu Met Ile
182               100                 105                 110
183 Met Ser Asp Glu Ala Val Glu Glu Arg Arg Ala Leu Ile Lys Arg Lys
184         115                 120                 125
185 Lys Ser Glu Arg Thr Gly Thr Gln Pro Leu Gly Val Gln Gly Leu Thr
186         130                 135                 140
187 Glu Glu Gln Arg Met Met Ile Arg Glu Leu Met Asp Ala Gln Met Lys
188 145                 150                 155                 160
189 Thr Phe Asp Thr Thr Phe Ser His Phe Lys Asn Phe Arg Leu Pro Gly
```

Input Set : A:\Sequence.txt
Output Set: N:\CRF3\05092000\I458366.raw

```
     190                   165                  170                    175
     191 Val Leu Ser Ser Gly Cys Glu Leu Pro Glu Pro Leu Gln Ala Pro Ser
     192            180               185                   190
     193 Arg Glu Glu Ala Ala Lys Trp Ser Gln Val Arg Lys Asp Leu Cys Ser
     194          195               200                   205
     195 Leu Lys Val Ser Leu Gln Ala Ala Gly Gly Gly Trp Gln Cys Leu Glu
     196       210               215                   220
W--> 197 Leu Gln Xaa Pro Ser Arg Gln Trp Arg Lys Glu Ile Phe Ser Leu Leu
     198 225                   230                  235                    240
     199 Pro His Met Ala Asp Met Ser Thr Tyr Met Phe Lys Gly Ile Ile Ser
     200                245               250                   255
     201 Phe Ala Lys Val Ile Ser Tyr Phe Arg Asp Leu Pro Ile Glu Asp Gln
     202             260               265                   270
     203 Ile Ser Leu Leu Lys Gly Ala Ala Phe Glu Leu Cys Gln Leu Arg Phe
     204          275               280                   285
     205 Asn Thr Val Phe Asn Ala Glu Thr Gly Thr Trp Glu Cys Gly Arg Leu
     206       290               295                   300
     207 Ser Tyr Cys Leu Glu Asp Thr Ala Gly Gly Phe Gln Gln Leu Leu Leu
     208 305                   310                  315                    320
     209 Glu Pro Met Leu Lys Phe His Tyr Met Leu Lys Lys Leu Gln Leu His
     210                325               330                   .335
     211 Glu Glu Glu Tyr Val Leu Met Gln Ala Ile Ser Leu Phe Ser Pro Asp
     212             340               345                   350
     213 Arg Pro Gly Val Leu Gln His Arg Val Val Asp Gln Leu Gln Glu Gln
     214          355               360                   365
     215 Phe Ala Ile Thr Leu Lys Ser Tyr Ile Glu Cys Asn Arg Pro Gln Pro
     216       370               375                   380
     217 Ala His Arg Phe Leu Phe Leu Lys Ile Met Ala Met Leu Thr Glu Leu
     218 385                   390                  395                    400
     219 Arg Ser Ile Asn Ala Gln His Thr Gln Arg Leu Leu Arg Ile Gln Asp
     220                405               410                   415
     221 Ile His Pro Phe Ala Thr Pro Leu Met Gln Glu Leu Phe Gly Ile Thr
     222             420               425                   430
     223 Gly Ser
     226 <210> SEQ ID NO: 3
     227 <211> LENGTH: 25
     228 <212> TYPE: DNA
     229 <213> ORGANISM: Artificial Sequence
     231 <220> FEATURE:
     232 <223> OTHER INFORMATION: putative SXR response element from the steroid
     233       hydoxylase, rCYP3A1
     235 <400> SEQUENCE: 3
     236 tagacagttc atgaagttca tctac                                           25
     238 <210> SEQ ID NO: 4
     239 <211> LENGTH: 25
     240 <212> TYPE: DNA
     241 <213> ORGANISM: Artificial Sequence
     243 <220> FEATURE:
     244 <223> OTHER INFORMATION: putative SXR response element from the steroid
```

```
          RAW SEQUENCE LISTING              DATE: 05/09/2000
          PATENT APPLICATION: US/09/458,366  TIME: 11:12:47

          Input Set : A:\Sequence.txt
          Output Set: N:\CRF3\05092000\I458366.raw
```

```
245        hydoxylase, rCYP3A2
247 <400> SEQUENCE: 4
248 taagcagttc ataaagttca tctac                          25
250 <210> SEQ ID NO: 5
251 <211> LENGTH: 25
252 <212> TYPE: DNA
253 <213> ORGANISM: Artificial Sequence
255 <220> FEATURE:
256 <223> OTHER INFORMATION: putative SXR response element from the steroid
257        hydoxylase, rUGT1A6
259 <400> SEQUENCE: 5
260 actgtagttc ataaagttca catgg                          25
262 <210> SEQ ID NO: 6
263 <211> LENGTH: 26
264 <212> TYPE: DNA
265 <213> ORGANISM: Artificial Sequence
267 <220> FEATURE:
268 <223> OTHER INFORMATION: putative SXR response element from the steroid
269        hydoxylase, rbCYP2C1
271 <400> SEQUENCE: 6
272 caatcagttc aacagggttc accaat                         26
274 <210> SEQ ID NO: 7
275 <211> LENGTH: 33
276 <212> TYPE: DNA
277 <213> ORGANISM: Artificial Sequence
279 <220> FEATURE:
280 <223> OTHER INFORMATION: putative SXR response element from the steroid
281        hydoxylase, rP450R
283 <400> SEQUENCE: 7
284 cacaggtgag ctgaggccag cagcaggtcg aaa                 33
286 <210> SEQ ID NO: 8
287 <211> LENGTH: 27
288 <212> TYPE: DNA
289 <213> ORGANISM: Artificial Sequence
291 <220> FEATURE:
292 <223> OTHER INFORMATION: putative SXR response element from the steroid
293        hydoxylase, rCYP2A1
295 <400> SEQUENCE: 8
296 gtgcaggttc aactggaggt caacatg                        27
298 <210> SEQ ID NO: 9
299 <211> LENGTH: 27
300 <212> TYPE: DNA
301 <213> ORGANISM: Artificial Sequence
303 <220> FEATURE:
304 <223> OTHER INFORMATION: putative SXR response element from the steroid
305        hydoxylase, rCYP2A2
307 <400> SEQUENCE: 9
308 gtgctggttc aactggaggt cagtatg                        27
310 <210> SEQ ID NO: 10
```

Please Note:
Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

VERIFICATION SUMMARY   DATE: 05/09/2000
PATENT APPLICATION: US/09/458,366   TIME: 11:12:48

Input Set : A:\Sequence.txt
Output Set: N:\CRF3\05092000\I458366.raw

```
L:99 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1
L:100 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1
L:197 M:258 W: Mandatory Feature missing, <221> not found for SEQ ID#:2
L:197 M:258 W: Mandatory Feature missing, <222> not found for SEQ ID#:2
L:197 M:340 W: (46) "n" or "Xaa" used: Feature required, for SEQ ID#:2
L:460 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:22
L:477 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:23
```